4 APPLICATIONS OF NEURAL NETWORKS IN MEDICINE AND BIOLOGICAL SCIENCES

Faramarz Valafar

4.1 INTRODUCTION

In this chapter, we will discuss applications of artificial neural networks (ANNs) in medicine and biological sciences. In particular, we will discuss ANN solutions to classical engineering problems of detection, estimation, extrapolation, interpolation, control, and pattern recognition as it pertains to these sciences. We will discuss some of these applications in detail to introduce the readers to typical problems that researchers face in the area.

Research in ANNs' applications as an alternative to classical engineering and mathematical techniques in medicine and biological sciences has intensified in recent years. Since the early 1990s, many applications of ANNs have replaced classical solutions to the engineering problems mentioned above. This is also true in medicine and biological sciences. [1 – 20] To discuss applications and accomplishments of ANNs in medicine and biological sciences, we will first introduce a few standard measures that will be used throughout this chapter to compare or report various results. These measures have been recommended and used to evaluate physicians and healthcare workers by various organizations, and therefore are good measures for evaluating the performance of any automated system that is designed to assist these healthcare professionals.

4.2. TERMINOLOGY AND STANDARD MEASURES

The American Heart Association (AHA) recommends the use of four measures to evaluate procedures for diagnosing CAD. [21] Since these measures are useful in other areas of diagnosis as well, we will be using them in evaluating most diagnostic systems.

$$sensitivity\ ..TPF = \frac{TP * 100}{TP + FN} \tag{4.1}$$

$$specificity\ ..TNF = \frac{TN * 100}{TN + FP} \tag{4.2}$$

$$PA = sensitivity * P(D) + specificity * [1 - P(D)] \tag{4.3}$$

VISIT…

$$PV = \frac{sensitivity * P(D)}{sensitivity * P(D) + (100 - specificity) * [1 - P(D)]} \tag{4.4}$$

Where *TP* stands for true positive, *FN* stands for false negative, *TN* stands for true negative, and *FP* stands for false positive. *Sensitivity*, or *true-positive fraction* (*TPF*), is the probability of a patient who is suffering from a disease to be diagnosed as such. *Specificity*, or *true-negative fraction* (*TNF*), is the probability that a healthy individual is diagnosed as such by a diagnosis mechanism for a specific disease. *PA* is the *predictive accuracy*, or the overall percentage of correct diagnosis. *PV* is the *predictive value* of a positive test, or the percentage of those who have the disease and have tested positive for it. *P(D)* is the *a priori* probability of a patient who is referred to the diagnosis procedure actually having cancer.

In addition to *TPF* and *TNF*, we define two other related values. *False-positive fraction* (*FPF*) is the probability of a healthy patient being incorrectly diagnosed as having a specific disease. And *false-negative fraction* (*FNF*) is the probability that a patient who is suffering from a disease will be incorrectly diagnosed as healthy. In this way, the following relations can be established:

$$FPF = 1 - TNF \tag{4.5}$$
$$FNF = 1 - TPF \tag{4.6}$$

To clarify the terminology and symbols, let us consider the following example.

Example 4.1:

Let us assume that 100 patients were referred to the mammography department for diagnosis of breast cancer. Let us further assume that of the 100 individuals, 38 actually had a cancerous tumor, and the remaining 62 either did not have any tumor or did not have one that was malignant (cancerous). Let us further assume that a diagnosis procedure (manually conducted by physicians, by an automated system, or by both) correctly diagnosed 32 of the 38 cancer sufferers as having breast cancer. It, however, misdiagnosed six of those as being cancer free. Let us also assume that the procedure correctly classified 58 of the 62 cancer-free patients as such, and misclassified the remaining 4 as having breast cancer. Finally, let us assume that on the average, 35 % of those who are referred to the mammography procedure actually have breast cancer.

In this example *TP = 32, FN = 6, TN = 58, FP = 4, and P(D) = 0.35.* Hence,

$$Sensitivity \equiv TPF = \frac{32 * 100}{32 + 6} = 84.21\% \Rightarrow FNF = 1 - 84.21 = 15.79\%,$$

$$Specificity = TNF = \frac{58 * 100}{58 + 4} = 93.55\% \Rightarrow FPF = 1 - 93.55 = 6.45\%,$$

$$PA = 84.21 * 0.35 + 93.55 * [1 - 0.35] = 90.28\%,$$

$$PV = \frac{84.21 * 0.35}{84.21 * 0.35 + (100 - 93.55) * [1 - 0.35]} = 87.55\%$$

In this example the overall system accuracy is 90.28 %, while the predictive value of a positive test is at 87.55 %.

Another commonly used measure of ANNs' performance that has found its way into the medical community (among others) is the receiver operating characteristic (ROC) curve. [22,23]. ANNs that perform pattern recognition or detection could be viewed as a receiver system (in the sense of a radar signal receiver) that receives a noisy signal and attempts to identify it. In the radar example, identification of the signal could mean classifying an aircraft as friend or foe. In medical decision-making, it usually means the diagnosis of a patient as healthy or sick. For simplicity, let us assume that the ANN has one output neuron. The following discussion can be expanded to cover multi output ANNs as well.

An important variable in the performance of the ANN is the threshold value θ of the output neuron. If $\theta=1$, all incoming signals in radar technology would be classified as noise. In medical technology, it would translate into having a negative diagnosis for all patients and, thus, categorizing them as healthy. If $\theta=0$, we would be classifying all patients as sick. In the first case, the probability of detection, or *TPF,* would have a value of zero, but so would the probability of false alarm, or *FPF*. In the second case, *TPF* would have a value of one, as would *FPF*. Neither of these receivers (detectors) would be desirable. The ROC curve gives an idea as to how the receiver would perform for all values of threshold (θ) between 0 and 1.

Let us assume that $f\left(z\middle|H_s\right)$ is the conditional probability density function of z, the activation level of the output neuron, given that the input to the network contains a signal (patient actually is carrying the disease) (hypothesis H_s). Similarly, $f\left(z\middle|H_n\right)$ is the conditional probability density function of z, given that the input to the network does not contain any signal (patient is actually healthy) (noise only, hypothesis H_n). A hypothetical example of these two density functions is shown in Figure 4.1.

With θ_0 being the alarm threshold, the probability of detection, or *TPF,* and the probability of positive error, or *FPF,* can then be calculated as follows:

$$TPF\left(\theta_0\right)=\int_{\theta_0}^{1} f\left(z\middle|H_s\right)dz \tag{4.7}$$

$$FPF\left(\theta_0\right)=\int_{\theta_0}^{1} f\left(z\middle|H_n\right)dz \tag{4.8}$$

Figure 4.2 shows the *TPF and FPF* graphs for the probability density functions of the hypothetical example shown in Figure 4.1.

Definition 4.1: The ROC curve of a system is the plot of that system's *TPF* curve versus its *FPF* curve. The operating variable of

the three curves is θ_0, the alarm threshold.

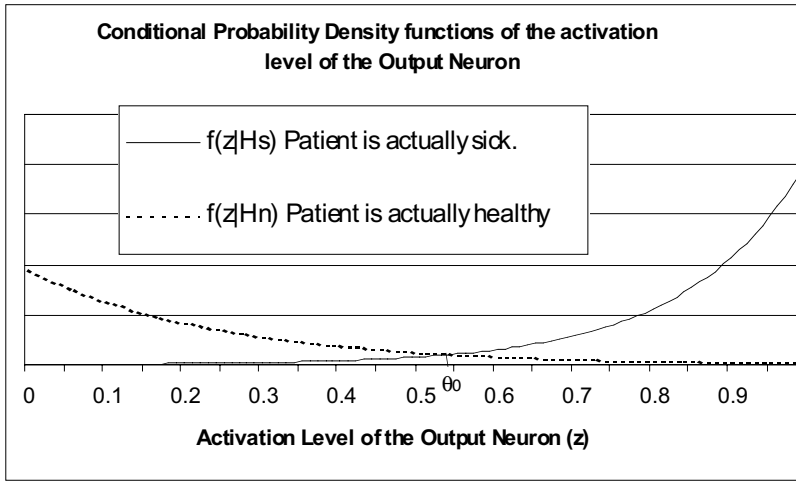

Figure 4.1: Hypothetical Conditional Probability Density Functions of the Activation Level z of the Output Neuron, Given that the Patient is Known to be Sick (H_s) or Healthy (H_n).

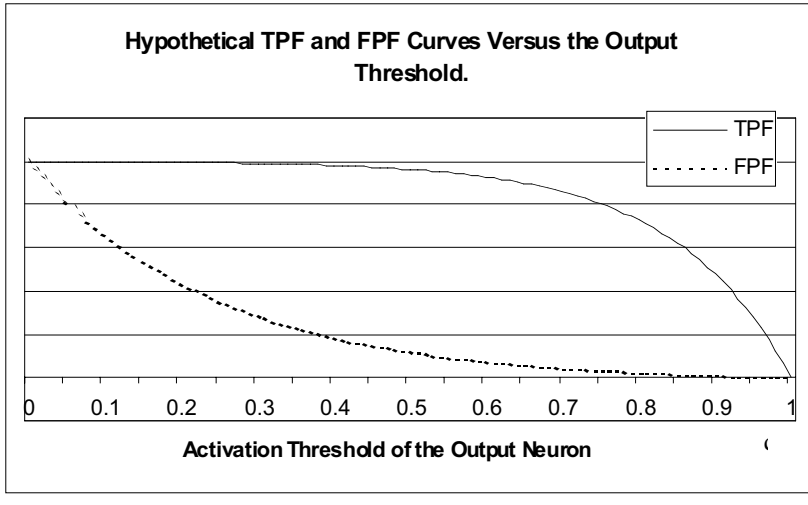

Figure 4.2: The True Positive Fraction and the False Positive Fraction Functions of the Hypothetical Example Shown in Figure 4.1, Plotted vs. the Alarm Threshold of the Output Neuron.

The ROC plot of the hypothetical example can now be plotted according to Definition 4.1. Figure 4.3 shows the ROC plot for the hypothetical example of

Figure 4.1. This figure also shows a worst case classifier (dashed line), and a theoretical best case classifier.

The ROC curve demonstrates an important property of any detection system: namely, that the probability of "true positive" is directly related to the probability of "false positive." They rise and fall together. The ideal classifier is one whose TPF is one for all values of FPF, including when $\theta_0 = 1$ and $FPF = 0$ (the red curve in Figure 4.3). The worst classifier is one that has no discrimination. A positive detection always has equal probability of being true or false. In other words, $TPF = FPF$ for all values of θ_0. This, in turn, would produce the dashed line ROC curve shown in Figure 4.3.

A consolidated measure that is a good representation of the overall quality of the receiver, and of the model used to build the receiver, is the area under the ROC curve. This area is commonly referred to as A_z. [23] This area varies between 0.5 (worst receiver) and 1 (best receiver). The area under the hypothetical ROC curve of the example in Figure 4.1 is 97.49 %. Furthermore, the best operating point of a receiver can be determined from the ROC curve by determining the point with a maximum distance from the diagonal line of the worst case classifier. In Figure 4.3, the label "Ideal Operating Point" shows this point of the hypothetical receiver. As can be seen from the figure, the best operating point of the hypothetical receiver has a TPF value of about 84 %, and FPF value of about 3.5 %.

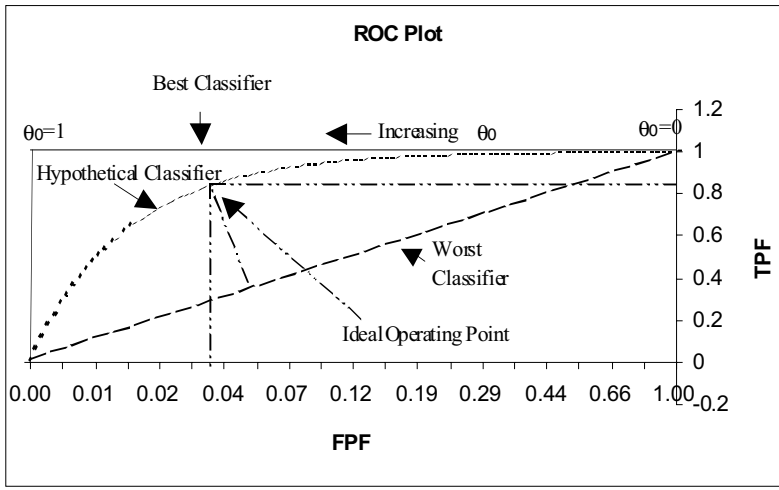

Figure 4.3: ROC Plots of a Hypothetical Receiver, a Theoretical Best Case Receiver, and a Worst Case Receiver.

The measures introduced in this section are particularly helpful in comparing the performance of most diagnosis procedures (systems) and therefore will be used in several parts of this chapter.

4.3 RECENT NEURAL NETWORK RESEARCH ACTIVITY IN MEDICINE AND BIOLOGICAL SCIENCES

ANNs have enjoyed success in various areas of medicine and biological sciences. ANNs have been successfully applied to areas such as radiology [16], cancer research, [12,14,24–29] biochemical spectrum analysis, [30] sleep disorder, [31] cardiac disease, [1,2,15,18,19] biochemistry of a disease, [4,5] HIV and AIDS, [5,29] epilepsy, [6,20] vision, [7] motor control, [8] lunge disease, [10,11] pathology and laboratory data analysis, [13,14] diagnosis decision support, [17,18,32] and many more.

In the following we present a brief summary of three research projects as examples of some of the most active research areas in ANNs' applications in medicine and biological sciences. These three applications are only meant to give an indication as to the breadth of the activity areas, and to demonstrate the typical problems (and give some ideas as to possible solutions) that researchers often face when dealing with real world data in the areas of medicine and biological sciences. It is also our hope that through these examples we can indicate the level of achievement of the ANN research community in various fields.

4.3.1 ANNs in Cancer Research

Pattern recognition using ANNs in cancer research is likely to be the most active area in terms of application of ANNs in medicine. ANNs have been used extensively in various roles in cancer research anywhere from tumor detection and analysis, [24,25,26] to the detection of biochemical changes in the body due to cancer, [29] to analysis of follow-up data in primary breast cancer, [27] to visualizing anticancer drug databases. [28] Among various types of cancer and detection methods, breast cancer diagnosis by the means of ANN classification of mammography images has been one of the most widely studied.

T.C.S.S. André and A. C. Roque [24] offer one of the most recent studies in this area. The authors have developed a medical decision support system using neural networks to aid in the diagnosis of breast cancer. This system uses digital mammogram images to classify a case as having one of three possible outcomes: suspicion of malignant breast cancer, suspicion of benign breast cancer, or no suspicion of breast cancer. André and Roque [33] used a staged (layered) neural network with a set of identical single layer networks as the input layer. These input layer networks used localized receptive fields without overlapping in the mammogram image. The hidden and the output layers of the network were each a single layer of perceptrons. The input layer was first trained, with regions taken from several mammograms, to become a feature extractor using the competitive learning algorithm. [33] The perceptron layers were then trained with the backpropagation learning algorithm.

The authors report a *TPF* of 0.75, and an *FPF* value of 0.06 for the optimum operating point of the ANN system described above. Furthermore, they report

an A_z value of 0.84 for their system. To put this value in perspective, it should be mentioned that A_z values typically fall in the 0.80 to 0.90 range for mammography analysis. In a similar study, Wu et al. [34] report an A_z value of 0.84 for a group of attending radiologists, and an A_z value of 0.80 for a group of resident radiologists.

Wu et al. [34] also conducted a similar study using a neural network trained with the backpropagation algorithm. They used a set of features of mammogram images that were selected by experienced radiologists as the input signal to the neural networks. In this case, they report an A_z value of 0.95 for textbook cases, and an A_z value of 0.89 for clinical cases.

4.3.2 ANN Biosignal Detection and Correction

Applications of signal detection techniques have been used in biological sciences to detect a single signal, or a group of signals, buried in various types of noise and nonrelevant biosignals for several decades. Applying pattern recognition techniques to spectroscopic data, for instance, has been used to help in structural elucidation of known molecules, and to significantly reduce the enormous duplicate work otherwise conducted in the area. Pattern recognition tools can therefore be employed to build search engines for spectral databases of various types of molecules.

An example of this can be seen in detecting the signature of one, or a group of complex carbohydrates in gas chromatography-electron impact mass spectroscopy (GC-EIMS), or nuclear magnetic resonance (NMR) spectra. [30]

Complex carbohydrates have been linked to biochemical functions of all cells, [35 – 37] such as cell recognition (e.g., initial steps in host pathogen and symbiotic relationships), intercellular adhesion (lectins and selectins), biosignal processes (oligosaccharins), developmental regulation, antibody binding, immune system modulation, and hormonal regulation. Consequently, complex carbohydrates, or the receptors that bind them, are also involved in many diseases, including autoimmune diseases, inflammatory diseases, and cancer. A tool to rapidly elucidate the chemical structures of complex carbohydrates can be instrumental in research to understand their biological functions. The presence of specific carbohydrates or their "uncommon" relatives, for instance, could be indicative of disease, the stage of a disease, or the presence of an antibody.

In this context, signal correction techniques could be used to correct the incoming biosignals and to compare them to a prerecorded clean library of signals. In this way, signal detection and correction techniques are used to discover and clean biosignals, and subsequently identify complex carbohydrates from which they originated. In this section, we discuss an artificial neural network solution to biochemical signal detection and identification, as well as biochemical signal correction for complex carbohydrates. [30]

Identification Of Complex Carbohydrate Structures from Their Spectral Signatures Using ANNs.

Structural and functional elucidation of complex carbohydrates is a key part of an increasing number of biomedical inquiries into these molecules. The structural determination of complex carbohydrates is the mandatory prerequisite to determining their functions. But the enormous chemical complexity and diversity of complex carbohydrates makes their structural elucidation a particularly challenging, lengthy task, and one that scientists would not wish to duplicate unnecessarily. Therefore, the primary need for the scientist faced with finding out the identity, chemical characteristics, and other attributes of a carbohydrate is to know whether that carbohydrate has already been analyzed by others and, if so, what is known about its chemistry, biology, and conformation. F. Valafar and H. Valafar [30] have developed a system for automated identification of complex carbohydrates using their chemical spectra that can provide this type of information.

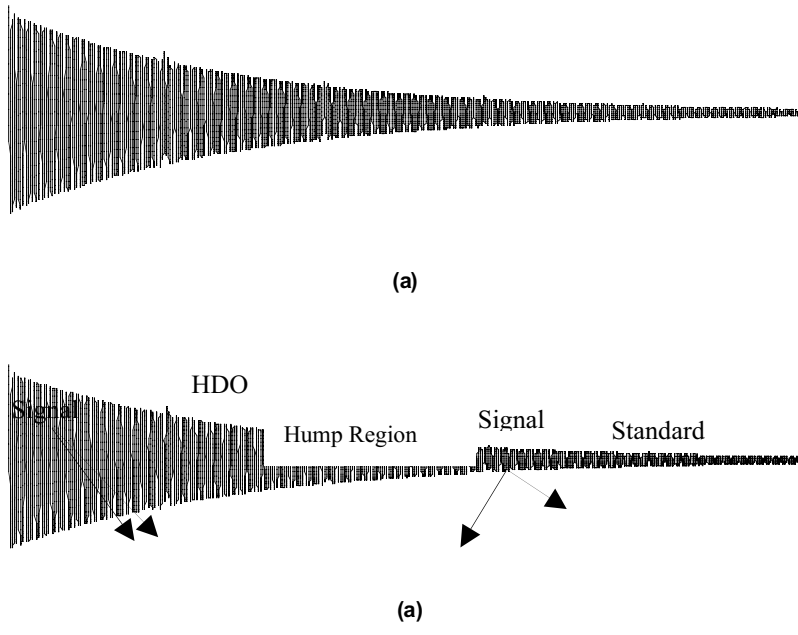

Figure 4.4 (a) A ^{1}H-NMR Time-Domain Signal of an *N*-linked Oligosaccharide. (b) The Fourier Transformed Frequency-Domain Spectrum of the Same Oligosaccharides.

In the following, we discuss Valafar's method in identifying complex carbohydrate structures from their ^{1}H-NMR spectra using artificial neural networks. In most classical signal processing methodology, the process of structural elucidation of a chemical compound from its ^{1}H-NMR spectrum first involves individual signal detection of elementary components (proton or ^{1}H signals). The second step in this process is the task of combining the detected individual signals in order to identify the structure of the carbohydrate in question. Valafar's use of ANNs in this process combines the two steps; the ANN performs both steps at the same time.

^{1}H-NMR spectra, in general, suffer from environmental, instrumental, and other types of variations that manifest themselves in a variety of aberrations. Low signal-to-noise ratio, [38 – 40] baseline drifts, [41 – 43] frequency shifts due to temperature variations, line broadening and negative peaks due to phasing problems, and malformed peaks (or peaks overlapped more than usual) due to inaccurate shimming are among the most common aberrations. Figure 4.4 demonstrates a clean ^{1}H-NMR spectrum of an *N*-linked complex carbohydrate.

As can be seen from Figure 4.4, large peaks not relevant to the structural elucidation of the complex carbohydrate usually dominate ^{1}H-NMR spectra of complex carbohydrates. These peaks include that of the solvent (heavy water in this case, HDO) and that of the standard. The proton signals (drifts) are typically in the order of 100 times weaker than the large peaks. Furthermore, most of these signals heavily overlap in the "hump" region of the spectrum, leaving the region unusable for structure elucidation.

For the purpose of automated identification of these spectra, elimination of the above mentioned aberrations becomes essential, as they can lead to erroneous identification. [41–45] A variety of signal processing techniques have been applied to "clean up" ^{1}H-NMR spectra. For instance, signal averaging[1] and apodization[2] have become standard ways of improving the signal-to-noise ratio. To correct baseline problems, a number of techniques have been used such as parametric modeling using *a priori* knowledge, [41,42] optimal associative memory (OAM), [42] spectral derivatives, [46] polynomial fitting, partial linear fitting, [47] and Bayesian analysis. [48] For peak detection (and solvent peak suppression), methods such as Bayesian analysis [48,49] and principal component analysis [50,51] can be mentioned. For signal-to-noise

[1] In signal averaging a spectrum is recorded several times. Each recorded signal is referred to as a "transient". The final spectrum is the arithmetic average of all the transients. The hope is that by using signal averaging the zero mean components of the noise present in the signal will be averaged out.[44]

[2] Apodization is a type of low (high) pass filtering performed in the time domain. Apodization is performed by speeding up or slowing down the rate of decay of time domain exponential functions. This is accomplished by multiplying the time domain signal by another function. This technique allows the improvement of the signal-to-noise ratio at the cost of the reduction in signal resolution (or vice versa).[44]

ratio problems, various types of filters (including adaptive filters such as matched filters [44,51]) in addition to standard apodization and signal averaging have also been used. A number of other mathematical techniques have also been introduced to address other specific types of aberrations encountered in ^{1}H-NMR spectra.

Although many of these signal processing techniques have enjoyed success, they remain solutions to specific types of aberrations. In order to produce sufficiently "clean" spectrum overall, one needs to use several of these methods to eliminate the aberrations present in a real spectrum. Furthermore, most of these techniques produce side effects that are magnified when improperly processed by a second signal processing algorithm, which can lead to false identification. Moreover, after the initial signal processing steps have been taken, the task of identifying the processed spectrum remains. This is not a trivial task as frequently the quality of the processed spectrum remains poor, requiring a sophisticated identification system.

Valafar and Valafar [30] have developed an artificial neural network system that addresses many of the above mentioned problems while identifying ^{1}H-NMR spectra of complex carbohydrates. Although the procedure still requires a minimal amount of preprocessing, it has significantly reduced the number of preprocessing steps while increasing the overall identification accuracy.

In this project, the authors developed an ANN system for a library of *N*-linked oligosaccharides, and one for xyloglucan oligosaccharides. While xyloglucans are plant cell wall oligosaccharides, the *N*-linked oligosaccharides are present in most animal biochemistry. Since the two systems used similar methods to develop an ANN identification system, we will only discuss here the development of the *N*-linked ANN identifier.

Preprocessing. Initial testing indicated that without preprocessing all selected methods for identification purposes would perform poorly. Therefore, it was decided to use some minimal preprocessing techniques to eliminate some aberrations before the identification stage. These preprocessing steps included baseline correction, high frequency noise reduction, and water and solvent peak elimination. These steps were respectively accomplished by a first derivative technique, a low-pass filter in the form of a specially designed averaging moving window, and a bin selection technique. The ANN eliminated the remaining aberrations in the process of identification. In other words, the ANN was able to learn during training to be insensitive to the remaining aberrations. Additionally, since each ^{1}H-NMR spectrum contained anywhere from 4K to 16K of data, an interpolation technique was used to normalize the length of all ^{1}H-NMR vectors to 5000. This would reduce (in most cases) the resolution of the spectrum to 2 points per Hertz, which is as low as Nyquist's theorem [51] would permit. The 5000-point vector covered the region between 1 and 5.5 ppm. The corrected spectra then were introduced to the ANN for training purposes.

Figure 4.5 shows the estimated *a posteriori* probability density functions [51] of the inter-[3] and intra-class[4] correlation coefficients between the raw (not processed) ^{1}H-NMR spectra of the *N*-linked data set[5] as defined by Bayes' theorem. [51] The required *a priori* density functions by Bayes' theorem were estimated using the nonparametric approach of Parzen density estimation. [52] Figure 4.6 shows the estimated *a posteriori* density functions of the preprocessed spectra from the same data set. As can be seen from the graphs, the overlap of the two density functions has been reduced from 56 to 43 %. This means that the "classical" signal preprocessing has simplified the identification task, and a Bayes' classifier, in combination with correlation coefficient analysis, now carries a 43 % uncertainty factor vs. the previous 56 %. Moreover, the probability density functions behave closer to expected (one large peak per density function, and smooth decay everywhere else in the function) after preprocessing.

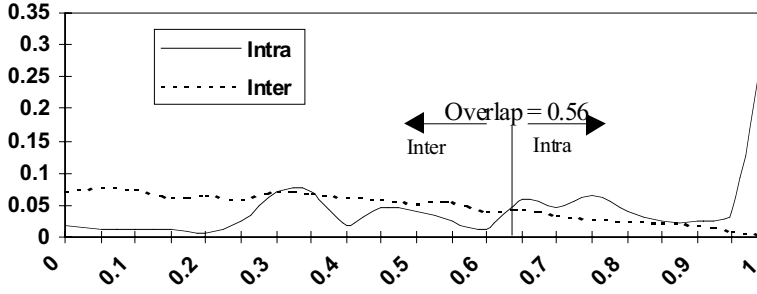

Figure 4.5 Estimated Distribution of Inter- and Intra-class Correlation Coefficients of Raw (Not Processed) ^{1}H-NMR Spectra of 109 ^{1}H-NMR Spectra of 23 *N*-linked Oligosaccharides.

ANN design

The authors used a two stage feedforward network with sigmoidal artificial neurons [33] in the hidden and output layers. The input layer of the network contained 5000 fan out neurons. The output layer contained 67 neurons

[3] By "class" we refer to the set of all spectra for a specific compound. In other words, each class in the xyloglucan experiment contained two spectra. In the *N*-linked database, 20 oligosaccharides were represented by five spectra, and the remaining three had three spectra, giving rise to five-member and three-member classes respectively. An "inter-class" correlation coefficient is the correlation coefficient between the spectra of two different oligosaccharides.

[4] "Intra-class" correlation coefficient is the correlation coefficient between two different spectra of the same oligosaccharide.

[5] The estimated Bayes' *a posteriori* distribution functions for the xyloglucan data set were similar to those shown here for the *N*-linked data set, and for space consideration are not shown here.

corresponding to the 67 oligosaccharides in the library. The number of the hidden neurons was empirically determined to be 27.

To develop the best performing ANN, several criteria were set forward: 1) the developed ANN was to have a very low FPF. In other words, if a spectrum of a complex carbohydrate was not present in the training library, the system should not try to find the closest match in the library. The outcome should be that the carbohydrate does not exist in the library; 2) the system needed to be tolerant of aberrations, and to be able to identify carbohydrates from its library even in the presence of relatively low signal-to-noise ratio. This translated into a high value for the area under the ROC curve, A_z. This also meant a high TPF value and 3) in the case of a mixture, the system was to indicate the carbohydrate of the highest ratio in the mixture.

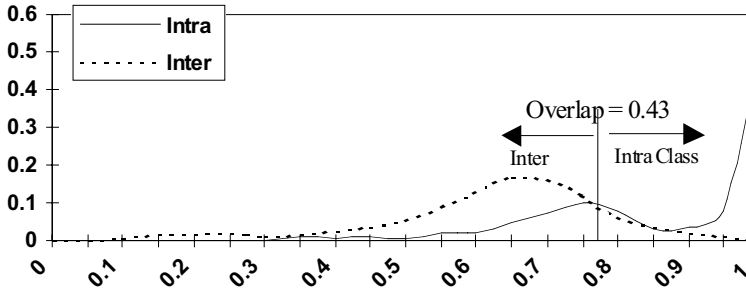

Figure 4.6 Histogram of the Correlation Coefficient Distribution of the Preprocessed ^{1}H-NMR Spectra of 109 ^{1}H-NMR Spectra of 23 *N*-linked Oligosaccharides.

With these goals in mind, a large number of training simulations were conducted. A large number of permutations were tried, namely, by varying the learning step size update policy, the number of hidden neurons, and the level of input noise. Valafar et al. dynamically manipulated the spectra during training by introducing some input noise in order to simulate the natural variability of these spectra. The noise simulated varying coupling constants due to temperature, line shape problems due to incorrect shimming, and minor baseline drifts. [53]

Table 4.1 shows the results of the best performing ANN in comparison with three other methods. The table shows the results of the experiments for both the *N*-linked and xyloglucan oligosaccharides.

Method A: Correlation coefficient analysis; Method B: Singular value decomposition; Method C: Correlation coefficient analysis and Bayesian classifiers; Method D: Backpropagation ANN.

The ANN system also showed less sensitivity to signal to noise degradation. Table 4.2 shows the degradation of identification accuracy of the four methods with increasing noise.

Table 4.1: Number of Correctly Identified *N*-linked and Xyloglucan Oligosaccharide Spectra (Total Number of Spectra is in Parentheses) by Four Different Identification Techniques after the Spectra Were Preprocessed as Described Above.

Method	*N*-linked Oligosaccharides		Xyloglucan Oligosaccharides	
	Training (67 spectra)	Testing (134 spectra)	Training (20 spectra)	Testing (20 spectra)
A	41	69	9	12
B	43	72	10	11
C	44	78	10	13
D	67	128	20	20

Table 4.2: Percentage Correct Identification of the Four Systems with Increasing Noise During Testing in the N-linked Oligosaccharide Database.

Method	Testing Noise Level				
	0%	5%	10%	15%	20%
A	51.49	41.86	37.21	34.88	27.91
B	53.73	46.51	39.53	34.88	25.58
C	58.21	53.49	46.51	37.21	32.55
D	95.52	95.35	81.40	62.79	39.53

4.3.3 Decision-making in Medical Treatment Strategies

Decision-making techniques can be used in medicine to solve various problems. Specifically, ANN pattern recognition engines have enjoyed significant success in medical decision-making. [1 – 20] Although, the ANN systems developed in this area demonstrate great potential benefit to the healthcare community, due to the numerous remaining challenges, the area remains one of the most active. To introduce the difficulties that researchers face in this area, we discuss here an ANN system designed to assist physicians in deciding on the best treatment strategy. Specifically, we will describe a research project conducted by H. Valafar et al. [32] to develop an ANN system to decide whether a beneficial, and yet at times harmful, medication (Hydroxyurea) should be prescribed in battling the symptoms of sickle cell anemia (SCA).

Predicting a sickle cell anemia patient's response to Hydroxyurea. Sickle cell anemia is a genetic disease mostly affecting African Americans in the US., although the disease is not limited to people from African origin worldwide. Treatment with Hydroxyurea (HU) partially alleviates disease symptoms in many patients with SCA.

Treatment with HU alleviates the clinical course in many patients with sickle cell anemia. [54] Most patients respond to HU with an increase in the fetal

hemoglobin (HbF) concentration of blood by either increasing the amount of HbF in their F-cells and/or by increasing the proportion of F-cells. The response to HU varies from patient to patient. If the magnitude of the HU-elicited increase in the %HbF (with respect to the total Hb) of the patient's blood could be predicted, "non-responders" could be identified. Although Hydroxyurea is effective for many patients, it is ineffective, and at times harmful, for others. Therefore, it is desirable to devise a tool with which physicians can predict, with a high percentage of accuracy, the outcome of the treatment before the medication is administered. Hence, the ultimate goal of the project is to *predict the response level of a given patient to Hydroxyurea, using only the pretreatment data of a patient*.

To develop such a system, the first question that needs to be answered is: *What data should be used for the prediction/decision-making task?* In this particular project, the authors relied on the expertise and experience of the physicians who were involved in sickle cell anemia research. The final set of data to be used for prediction contained the results of a standard blood test, in addition to some genetic information. A detailed list of the parameters that were used can be seen in Table 4.3.

Selection of the parameters listed in Table 4.3 was based solely on educated guesses on the physicians' parts (such as the genetic information), and some earlier simple statistical analysis of various data. Therefore, it could be expected that some of the 23 parameters might not be relevant to the problem at hand. It is also quite possible that not all relevant parameters are included in the study.

Data preprocessing. Many medical databases, especially those that go years into patients' past history and treatment, are in printed or written form. The first step in this research was to create an electronic database usable by the modeling team. This process was accomplished at the Medical College of Georgia. All patients' data were entered into a widely available spreadsheet. These data then were sent to the modeling team for analysis.

Soon after the first round of analysis was completed, the following problems were observed:

1) *Missing data.* A quick look at the data revealed that much of the data was missing. For instance, if the patient was feeling well in that particular month, certain measurements (tests) were not conducted. Furthermore, there were instances where the patient simply did not show up for follow-up tests because he/she was feeling okay. In some instances, the paperwork containing the data for the early stages of the treatment was misplaced and lost. There were two types of missing data in our databases. In some instances, certain variables (pieces of data) were missing from a monthly record. In others, an entire monthly record was missing.
2) *Incorrect data.* Simple statistical correlation analysis revealed that there were some severe outliers. Most of these were traced back to human error. But there were also data that simply were off the chart, but not traceable to

any human error. All the human errors were corrected. However, the nontraceable extreme outliers were excluded from the study.

3) *Invalid or corrupt data.* In some cases, there were patients who became pregnant against the doctor's advice, or underwent a blood transfusion due to other complications in the middle of the treatment period. The data of such patients were excluded from the study as the effects of such events on a patient's blood chemistry and his or her ability to respond to Hydroxyurea was unclear.

Table 4.3 A Description of the 23 Parameters for Which Data was Obtained from the Patients. From H. Valafar, et al., [32].

Parameter	Description	Units
Age	Age of patient at the time of analysis	Days
Sex	Male/Female	F=1, M=2
NAGG	α Globin gene number	None
BAN	Number of BAN haplotypes	1,2, or 3* None
BEN	Number of BEN haplotypes	1,2, or 3* None
CAM	Number of CAM haplotypes	1,2, or 3* None
SEN	Number of SEN haplotypes	1,2, or 3* None
WGT	Weight of patient	Kg
%HbF	Fetal hemoglobin, as % total hemoglobin	None
HbF	Fetal hemoglobin, absolute value	g/L of blood
Hb	Total hemoglobin concentration	g/dL of blood
RBC	Red blood cell count	x 10^{12} / Liter
PCV	Packed cell volume (hematocrit)	Liter / Liter
RDW	% Variation in the size of red cells	None
Retic	Reticulocytes	x 10^{3}
MCV	Mean cell (erythrocyte) volume	Femtoliters
MCH	Mean cell hemoglobin	Picograms
WBC	White cell count	x 10^{9} / Liter
Polys	Polymorphonuclear leukocytes	x 10^{9} / Liter
Plats	Platelet count	x 10^{9} / Liter
Bili	Bilirubin concentration in blood	mg / dL
NRBC	Nucleated red blood cells seen in peripheral blood	Number per WBC
Duration	Duration of treatment a patient received to arrive at the maximum %HbF level	Days

*The actual values were 0,1,or 2, but 0 could not be used (see last paragraph under ANN Analyses).

Problem definition. Further problems arose as the team prepared for the first round of modeling experiments. One of the more fundamental problems, and often one that is usually difficult to solve in medical decision-making problems, was with the definition of the problem (problem statement). After further close examination of the data, it was realized that the definition of the problem was inadequate and that the experiments were destined to either fail, or to produce results that were medically useless. The original statement of the problem was as follows: "Develop a system that can accurately distinguish positive responders from the nonresponders using pretreatment data." Furthermore, a "positive responder" was defined to be "a patient whose initial percentage HbF (%HbF) doubles at some point during the treatment."

After looking at the data, it was soon realized that while this definition may work for patients whose initial %HbF is, say 7%, or higher, it does not work so well for patients whose initial %HbF is 1% or 2%. In other words, while Hydroxyurea treatment might increase a patient's initial %HbF value from 1% to 2% at some point during treatment, it is not very likely that he/she would experience any benefits (reduced number of hospital visits, or reduced severity of symptoms) as a result of this minor increase. This meant that even in the bestcase scenario that a system with 100% accuracy (in separating the patients who can double their initial %HbF from those who cannot due to Hydroxyurea) was developed, its results would be clinically meaningless. This is because doubling the %Hbf value does not translate into reduced symptoms or hospital visits for many or all patients. A new definition for a "positive response" had to be devised.

After extensive study of published articles on Hydroxyurea and its alleviation of symptoms, two possible definitions were suggested:

1) *Dynamic patient threshold*. It was suggested that each patient has a different level of % HbF, beyond which his/her symptoms begin to taper off. A patient would be categorized as a positive responder if his/her %HbF level increased above this dynamic threshold as a result of the treatment. This dynamic level needs to be calculated or estimated for each patient via some type of computational means. Although this measure is probably the more accurate measure of positive response, it was soon realized that in order to estimate accurately each patient's threshold, one would need to have the response model in hand. Since the response model was the final goal of the project, this definition seemed impractical and was therefore abandoned.
2) *Static threshold*. The team agreed that the next best definition was that of a static threshold across all patients. This threshold was determined by consulting existing publications and the collaborators at MCG. All these sources seemed to agree that most patients experienced some type of relief of symptoms when their %Hbf rose about 15%. [55,57] Hence, if a patient's HbF concentration rose above 15% of total Hb during treatment, he/she was categorized as a positive responder, and all others as nonresponders. Three patients were excluded from this study, as their

initial % HbF was higher than 15. This threshold divides the final 83 patients included in the study into 58% responders and 42% non-responders.

Missing Data. The problem of missing data arises in medicine quite often. The most common causes of missing data are 1) patients who do not come into clinics for further tests when they start feeling better or, if they do come in, the nurses and the physicians who record the data are not as motivated to record all available information; and 2) data are commonly recorded on paper and, therefore, sometimes are misplaced and/or lost. While these are the two main causes of missing data, there are others that need not be mentioned here.

In general, regardless of the reason for missing data, the missing data can be categorized into two classes: 1) missing record: in some instances, the data for an entire record are missing. A common cause of this type of missing data in the case of SCA is due to patients who do not report to the clinic for their monthly tests when they experience some relief in their symptoms. In such cases, no data for that month are available for the patient; and 2) missing data points: In some instances, specific parameters in each record are missing. An example of this in the case of SCA would be when a patient who is feeling better reports to the clinic for a monthly test. In some such cases, not all the tests are conducted, or properly recorded. Human error is also a common source of this type of missing data.

The first type of missing data did not cause many problems in our experiments. This is because only the initial parameters of the patient (from before the beginning of the treatment) were used and the highest level of percentage HbF during treatment to train the artificial neural network. For this reason, missing intermittent data were not harmful to our experiments, except in cases when the highest percentage HbF was also missing. In the cases where the highest percentage HbF value was missing, all data of that patient were excluded from the study.

The second type of missing data could be potentially much more problematic, as it is much more likely for the value of some parameters to be missing at the initial recording before the beginning of the treatment. Since the initial values are vital information, all patients who were missing more than two initial parameters were dropped from the study. The patients whose data were missing one or two initial parameters were kept in the study as long as the level of initial percentage HbF was not missing. To fill in the missing parameters, some experiments were conducted with a few extrapolation algorithms. However, it was discovered that the best way to deal with the few missing parameters was to fill them in with zeroes. This is simply because Delta rule [33] and backpropagation algorithms were used to train our neural networks, and, as can easily be determined from weight update formulas, when the input parameter is zero, no learning is conducted in the first stage of the network. This was the best way to make use of the data without presenting the network with erroneous data.

Compliance. Compliance is one of the biggest problems in medical research. The simple cause of it is that some patients stop taking the medication, or at least reduce the dosage without instructions from the physician when they start feeling better. This can lead to corrupt data (for our purposes), as a patient could be falsely identified as a nonresponder. This was the case in our study. Our initial systems suffered from a relatively high FNF. From formula 4.6, it can easily be seen that this causes TPF to be reduced, and therefore A_z, the area under the ROC curve, to be lower than expected. As a result, it could lead to the false conclusion that the identification technique or system architecture is inadequate, while the source of the problem really lies in the data.

In the case of many medications, compliance can be measured by the variation in one or many biochemical parameters. This was the case with HU and SCA patients. One of HU's side effects is that it increases the volume of red blood cells. [58] Among the final 83 patients who were all categorized as compliant and were included in this study, the mean cell volume increased by an average of 22% as a result of HU treatment. This is in line with other studies. [55,56,58,59] The variable mean cell volume (MCV) is thus a good measure of compliance. This variable was analyzed for each patient. It was decided that six patients were not compliant and so their data were excluded from the study.

Figure 4.7 shows the bin distribution function of MCV before and after HU treatment. As can be observed, the distribution has clearly moved to higher values after the treatment and has a higher mean.

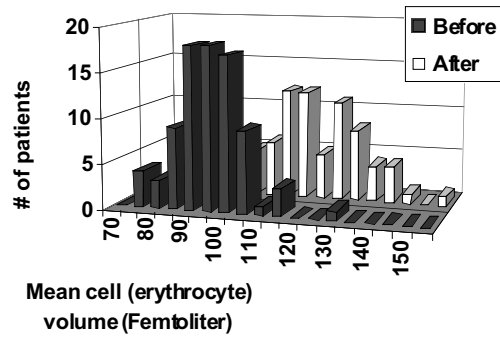

Figure 4.7: Distribution of Average Volume of the Red Blood Cells of 83 Sickle Cell Patients before and after Treatment with HU. From H. Valafar, et al., [32].

Neural network prediction model. An ANN using 23 input neurons, 4 hidden neurons, and 1 output neuron was used for the 15% threshold experiment. This neural network produced an output value higher than 0.5 if the patient was predicted to be a responder, and an output of less than 0.5 if the patient was predicted to be a nonresponder.

The threshold experiment was designed to eliminate the possibility that the ANN could simply "memorize" the values of the parameters of each patient. This was accomplished by training ANNs with the parameter values of 82 of the patients, and then using the values of the patient whose parameters had not been seen by the ANN, to test the ANN. This procedure was repeated 83 times and each time an ANN was trained. (A different patient was left out of the training each time) The result of this experiment is presented in Figure 4.8. Seventy patients were correctly classified as responders or nonresponders while 13 were misclassified. Thus, 84% of the responses were predicted correctly. This experiment was repeated five times with, on average, 86.6 correct predictions with a standard deviation of +/-2.0.

Variable selection. Researchers in the medical fields are also frequently faced with the problem of variable selection. In most cases, there is not enough information to select the relevant variables for a certain modeling/pattern recognition problem in medicine. Also, one of the reasons that researchers seek a mathematical model for a disease is to use it to determine the relevant variables. This information can be extremely helpful in understanding how the disease works, develops in the body, or is fought against by the body's immune system. In the latter case, if the immune system is failing to effectively fight the disease, information about relevant variables could lead to new medications that either help the body in eliminating the disease, or at least reduce its symptoms (e.g., the case of sickle cell anemia).

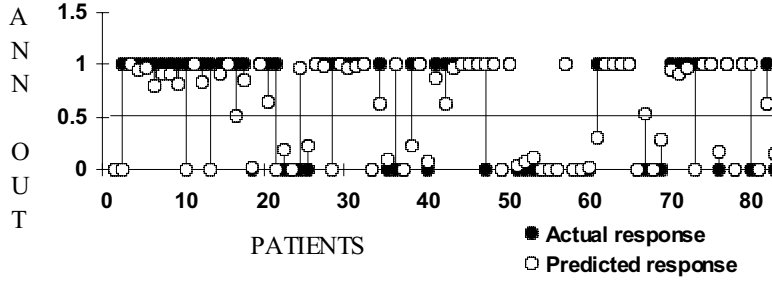

Figure 4.8: The Prediction by ANNs of Which Patients Would Respond to HU by an Increase in Their HbF Concentration to the Point Where it Accounts for 15% or More of Their Total Hb. ANNs Were Trained with the Values of the Parameters of 82 Patients and then Tested with the Values of the Parameters of the Patient that Had Not Been Used to Train the ANN. This Procedure Was repeated 83 times, each time Leaving Out a Different Patient and Training the ANN With the Data from the Other 82 Patients to Give the Values in the Figure. Patients Whose HbF Concentration Did Not Reach 15% of the Total Hb Should have generated an ANN "output" of less than 0.5, while patients whose HbF Concentration Exceeded 15% of the Total Hb Should Have Generated an ANN Output of More Than 0.5. From H. Valafar, et al., [32].

Valafar et al. designed their variable selection experiments in the SCA's case to identify which of the 23 parameters are most important or influential in assisting ANNs to predict those patients that will respond to HU treatment. Determining the importance of each of the 23 parameters was accomplished by employing two different methods. The first method consisted of a recursive elimination process in which a different set of parameters was taken out of the training set. The ANNs were trained with the values of the remaining parameters. The software measures the degradation of performance due to the missing parameters. This experiment is an exhaustive elimination process in which the removal of every combination of parameters (2^{23}-1=8,388,607 combinations) is evaluated. The degradation (or importance) of the parameters observed is the averages of ten experiments (two different seeds for the random number generator, and five runs per seed). The final effect of removing each set of parameters is calculated by averaging the performance degradation of the ten ANNs trained without that set of parameters.

The second method of parameter selection is an adaptive technique that takes into effect the synaptic connection strengths of each variable. This algorithm is initiated by setting equal values for each parameter. During the course of training, these values are updated to reflect the strength of the synaptic connection(s) associated with each parameter. This, in turn, is an indication of the contribution of each parameter towards the discovery of the correct answer. Thus, at the end of the training the contribution of each parameter reveals its importance in the solution of the problem. Each training session was repeated five times to eliminate any random behavior of the system.

Although the above two methods are distinctly different methods for parameter selection, both algorithms produced similar results in extracting the relevant parameters. For this reason, we will only discuss the results of the second method from this point forward.

The 23 parameters and their scores, which are proportional to their contributions in predicting the response to HU treatment, are listed in Table 4.4. This table contains the averaged data for over five different training sessions. The lack of any particularly influential contributors indicates that no one parameter contains the information needed to predict the response to HU. Therefore, based on the given contributions, it is reasonable to assume that the information needed for a successful classification is distributed among a number of parameters, perhaps even a fairly large number of parameters.

The ANNs whose testing results are shown in Figure 4.8 used the values of all 23 parameters. A separate experiment was carried out to determine if the values of just ten of the twenty-three parameters listed in the previous section could be used while maintaining the ANN's full ability to identify responders and non-responders. This experiment used the top ten parameters listed in Table 4.4. The ability to eliminate unnecessary parameters has the potential for reducing the problem size by more than 50%, and might assist in elucidating the mechanisms by which ANNs function.

Table 4.4 The Effectiveness of Each of the 23 Parameters to Assist ANNs in Predicting the Response of Patients to HU Treatment. From H. Valafar, et al., [32].

Parameter	Score
Duration	0.083
RDW	0.063
WBC	0.059
Plats	0.053
MCV	0.053
Polys	0.052
WGT	0.050
SSEN	0.045
Retic	0.043
Sex	0.042
SCAM	0.041
NAGG	0.041
Hb	0.040
SBAN	0.040
MCH	0.035
RBC	0.034
SBEN	0.034
Bili	0.034
Age	0.033
HbF	0.032
%HbF	0.031
PCV	0.031
SNBRC	0.030

The ANN trained only with the 10 selected variables had remarkably similar results to the one trained with all 23 variables. Except for 2 of the 83 patients, the results of the 2 networks were very similar. The network trained with ten variables produced outputs that were more clearly defined. The mean of the probability density function of the output z of the smaller network was higher for positive responders, and lower for nonresponders. By the same token, the standard deviation of both curves was smaller than those of the larger network. Furthermore, the two patients whose classification changed by using the smaller network were both marginally classified by the larger network. One was correctly classified as a responder, and one incorrectly as a nonresponder. With the smaller network, the first patient was incorrectly classified as a nonresponder; the second patient was correctly classified as a responder. Therefore, the TPF, FPF, and the ROC curves remained identical for both networks.

4.4 SUMMARY

Artificial neural networks have distinct features that can be advantageous in modeling natural phenomena in biology and medicine. Applications of ANNs in these fields are sure to help unravel some of the mysteries in various diseases and biological processes. In the SCA case, the ANN developed for the variable selection process helped pinpoint the parameters that possibly play an important role in understanding the works of SCA. This could lead to a significant increase in the life expectance of SCA sufferers.

Research in applications of ANNs in medicine and biological sciences currently remains strong. With more systematic data collection routines implemented in healthcare facilities, systems such as the ones described in this chapter are sure to find their way into doctors' offices and hospital laboratories.

REFERENCES

1. Akay, M., Akay, Y.M., Welkowitz, W., Semmlow, J.L., and Kostis, J.B., Noninvasive Detection of Coronary Artery Disease Using Neural Networks, *Proc. of the Ann. Conf. on Eng. in Med. and Biol.,* 13(3), 1434 – 1435, Oct 31 – Nov 3, 1991.
2. Akay, M., Noninvasive Diagnosis of Coronary Artery Disease Using a Neural Network Algorithm, *Biol. Cybern.,* 67: 361 – 367, 1992.
3. Alpsan, D., Auditory Evoked Potential Classification by Unsupervised ART 2-A and Supervised Fuzzy ARTMAP Networks, *Int. Conf. on Neural Networks (ICNN '94)*, IEEE, Orlando, FL, 3512 – 3515, June 26 – July 2, 1994.
4. Andrea, T.A. and Kalayeh, H., Applications of Neural Networks: Quantitative Structure-Activity Relationships of Dihydrofolate Reductase Inhibitors, *J. Med. Chem.,* 34:2824 – 2836, 1991.
5. Andreassen, H., Bohr, H., Bohr, J., Brunak, S., Bugge, T., Cotterill, R.M.J., Jacobsen, C., Kusk, P., and Lautrap, B. Analysis of Secondary Structure of the Human Immunodeficiency Virus Proteins by Computer Modelling Based on Neural Network Methods, *J. Acquired Immune Deficiency Syndrome,* 3, 615, 1990.
6. Apolloni, B., Avanzini, G., Cesa-Bianchi, N., and Ronchini, G., Diagnosis of Epilepsy via Backpropagation, *Proc. of the 1990 Int. Joint Conf. on Neural Networks*, Washington, DC, 2, 571 – 574, 1990.
7. Armentrout, S.L., Reggia, J.A., and Weinrich, M., A Neural Model of Cortical Map Reorganization Following a Focal Lesion, *Artif. Intelligence in Med.,* 6(5), Oct 1994.
8. Armstrong, W.W., Stein, B.A., Kostov, R., Thomas, M., Baudin, P., Gervais, P., and Popovic, D., Application of Adaptive Logic Networks and Dynamics to Study and Control of Human Movement, *Proc. of the Second Int. Symp. on 3D Anal. of Human Movement*, Poitiers, France, 81 – 84, June 30 – July 3, 1993.

9. Armstrong, W.W., Kostov, A., Stein, R.B., and Thomas, M.M., Adaptive Logic Networks in Rehabilitation of Persons with Incomplete Spinal Cord Injury, *Workshop on Environmental and Energy Applications of Neural Networks*, Richland, WA, Pacific Northwest National Laboratory, March 30 – 31, 1995.
10. Asada, N., Doi, K., MacMahon, H., Montner, S.M., Giger, M.L., Abe, C., and Wu, Y., Potential Usefulness of an Artificial Neural Network for Differential Diagnosis of Interstitial Lung Diseases: pilot study, *Radiology,* 177, Vol. 3, 857–60, December, 1990.
11. Asada, N., Doi, K., MacMahon, H., Montner, S., Giger, M.L., Abe, C., and Wu, Y., Neural Network Approach for Differential Diagnosis of Interstitial Lung Diseases, *Proc. SPIE (Medical Imaging IV)*, 1233: 45 – 50, 1990.
12. Ashenayi, K., Hu, Y., Veltri, R., Hurst, R., and Bonner, B., Neural Network Based Cancer Cell Classification, *Proc. of the World Congress on Neural Networks*, San Diego, CA, 1, 416 – 421 June 5 – 9, 1994.
13. Astion, M.L. and Wilding, P., The Application of Backpropagation Neural Networks to Problems in Pathology and Laboratory Medicine, *Arch. Pathol. Lab. Med.*, 116:995 – 1001, 1992.
14. Astion, M.L. and Wilding, P., Application of Neural Networks to the Interpretation of Laboratory Data in Cancer Diagnosis, *Clin. Chem.* (US) 38, 34 – 38, 1992.
15. Avanzolini, G., Barbini, P., and Gnudi, G. Unsupervised Learning and Discriminant Analysis Applied to Identification of High Risk Postoperative Cardiac Patients, *Int. J. Bio-Med. Comput.,* 25, 207 – 221, 1990.
16. Barski, L.L., Gaborski, R.S., and Anderson, P.G., A Neural Network Approach to the Histogram Segmentation of Digital Radiographic Images, *Intell. Eng. Sys. Through Artif. Neural Networks*, Dagli, Burke, Fernandez, and Ghosh, (eds.), 3, 375 – 380, ASME Press, NY, 1993.
17. Bartels, P.H., Thompson, D., and Weber, J.E., Diagnostic Decision Support by Inference Networks, *In Vivo,* 7, 379 – 385, 1993.
18. Baxt, W.G., Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making: the Diagnosis of Acute Coronary Occlusion, *Neural Computation,* 2, 480 – 489, 1990.
19. Baxt, W.G., Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction, *Ann. of Intern. Med.,* 115, 843 – 848, 1991.
20. Echauz, J. and Vachtsevanos, G., Neural Network Detection of Antiepileptic Drugs from a Single EEG Trace, *Proc. of the IEEE Electro/94 Int. Conf.*, 346 – 351, Boston, MA, May 10 – 12, 1994.
21. Gibbons, R.J., Balady, G.J., Beasley, J.W., Bricker, J.T., Duvernoy, W.F., Froelicher, V.F., Mark, D.B., Marwick, T.H., McCallister, B.D., Thompson, P.D. Jr., Winters, W.L., Yanowitz, F.G., Ritchie, J.L.,

Gibbons, R.J., Cheitlin, M.D., Eagle, K.A., Gardner, T.J., Garson. A. Jr., Lewis, R.P., O'Rourke, R.A., and Ryan, T.J., ACC/AHA Guidelines for Exercise Testing, A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing), *J. of the Am. Coll. of Cardiol.*, 30(1), 260 – 311, July, 1997.

22. Goodenough, D.J., Rossmann, K., and Lusted, L.B., Radiographic Applications of Receiver Operating Characteristic (ROC) Curves, *Radiology*, 110, 89 – 95, 1974.
23. Hanely, J.A. and McNeil, B.J., The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve, *Radiology,* 143, 29 – 36, 1982.
24. André, T.C.S.S. and Roque, A. C., A Neural Network System for the Diagnosis of Breast Cancer, *Proc. of the Int. Conf. on Math. and Eng. Techniques in Med. and Biol. Sci. 2000* (METMBS'00), Las Vegas, NV, 1, 1 – 6, June 26 – 29.
25. Rodrigues, R.G.S., Pela, C.A., and Roque, A.C., Tomographic Image Reconstruction Using Neural Networks, *FFCLRP*, Brazil, V1 27 – 33.
26. Chen, D., Chang, R.F., and Huang, Y.L., Breast Cancer Diagnosis Using Self-Organizing Map for Sonography, *Ultrasound. Med. Biol*., 26(3), 405 – 11, March, 2000.
27. Harbeck, N., Kates, R., Ulm, K., Graeff, H., Schmitt, M., Neural Network Analysis of Follow-Up Data in Primary Breast Cancer, *Int. J. Biol. Markers*, 15 Vol. 1, 116 – 22, January – March, 2000.
28. Shi, L.M., Fan, Y., Lee, J.K., Waltham, M., Andrews, D.T., Scherf, U., Paull, K.D., Weinstein, J.N., Mining and Visualizing Large Anticancer Drug Discovery Databases, *J. Chem. Inf. Comput. Sci*, 40, Vol. 2, 367 – 79, March – April, 2000.
29. Cherniak, R., Valafar, H., Morris, L.C., and Valafar, F., *Cryptococcus neoformans* Chemotyping by Quantitative Analysis of ^{1}H-NMR Spectra of Glucuronoxylomannans Using a Computer Based Artificial Neural Network, *J. of Clin. and Diag. Lab. Immunol.*, 5(2),146 – 159, March, 1998.
30. Valafar, F. and Valafar, H., CCRC-Net: An Internet-Based Spectral Database for Complex Carbohydrates, Using Artificial Neural Networks Search Engines, *Trends in Anal. Chem*., 18, 508 – 512, 1999.
31. Guimaraes, G., The Discovery of Sleep Apnea with Unsupervised Neural Networks, *Int. Conf. on Math. and Eng. Techniques in Med. and Biol. Sci. (METMBS'2000)*, 1, 361 – 367, Las Vegas, NV, June 26 – 29, 2000.
32. Valafar, H., Valafar, F., Darvill, A., Albersheim, P., Kutlar, A., Woods, C., and Hardin, J., Predicting the effectiveness of Hydroxyurea in Individual Sickle Cell Anemia Patients, *J. of Artif. Intell. in Med.,* 18 (2), 133 – 148, February, 2000.

33. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, NJ, 1999.
34. Wu, Y., Giger, M.L., Doi, K., Vyborny, C.J., Schmidt, R.A., and Metz, C.E., Artificial Neural Networks in Mammography: Application to Decision Making in the Diagnosis of Breast Cancer, *Radiology,* 187 Vol. 1, 81 – 7, April, 1993.
35. Varki, A., Biological Roles of Oligosaccharides: All the Theories are Correct, *Glycobiology,* 3, 97 – 130, 1993.
36. Goochee, C.F., Gramer, M.J., Andersen, D.C., Bahr, J.B., and Rasmussen, J.R., The Oligosaccharides of Glycoproteins: Factors Affecting Their Synthesis and Their Influence on Glycoprotein Properties, *Frontiers in Bioprocessing II.* (Todd, P., Sikdar, K., and Bier, M., eds.) 199 – 240, American Chemical Society, Washington, D.C., 1992.
37. Cook, G.M.W., Glycobiology of the Cell Surface: the Emergence of Sugars as an Important Feature of the Cell Periphery, *Glycobiology,* 5, 449 – 461, 1995.
38. Van Huffel, S., Enhanced Resolution Based on Minimum Variance Estimation and Exponential Data Modeling, *Signal Processing*, 33, 333 – 355, 1993.
39. Van den Boogaart, A., Howe, F.A., Rodrigues, L.M., Stubbs, M., Griffiths, J.R., *In Vivo* ^{31}P MRS: Absolute Concentrations, Signal-to-Noise and Prior Knowledge, *NMR in Biomed.*, 8, 87 – 93, 1995.
40. Angelidis, P.A., Spectrum Estimation and the Fourier Transform in Imaging and Spectroscopy, *Concepts Magn. Resonance*, 8 Vol. 5, 339 – 381, 1996.
41. Blumler, P., Greferath, M., Blumich, B., and Spiess, H.W., NMR Imaging of Objects Containing Similar Substructures, *Magn. Resonance*, Series A 103, 142 – 150, 1993.
42. Wabuyele, B.W. and Harrington, P., Optimal Associative Memory for Background Correction of Spectra, *Anal. Chem.*, 66, 2047 – 2051, 1994.
43. Wabuyele, B. W. and Harrington, P., Quantitative Comparison of Bidirectional Optimal Associative Memories for Background Prediction of Spectra, *Chemometrics and Intelligent Lab. Sys.*, 29, 51 – 61, 1995.
44. Angelidis, P. A., Spectrum Estimation and the Fourier Transform in Imaging and Spectroscopy, *Concepts Magn. Resonance*, 8(5), 339 – 381, 1996.
45. Goodacre, R., Timmins, E.M., Jones, A., Kell, D.B., Maddock, J., Heginbothom, M., Magee J. T., On Mass Spectrometer Instrument Standardization and Interlaboratory Calibration Transfer Using Neural Networks, *Analytica Chemica Acta*, 348, 511 – 532, 1997.

46. Gerow, D.D. and Rutan, S.C., Background Subtraction for Fluorescence Detection in Thin–layer Chromatography with Derivative Spectrometry and the Adaptive Kalman Filter, *Analytica Chemica Acta*, 184, 53, 1986.
47. Yu, K. M. and Jones, M.C., Local Linear Quantile Regression., *J. Am. Statistical Assoc.,* 93(441): 228 – 237, March, 1998.
48. Whittenburg, S., Baseline Roll Removal in NMR Spectra Using Bayesian Analysis, *Spectroscopy Letters*, 28(8), 1275 – 1279, 1995.
49. Whittenburg, S., Solvent Peak Removal in NMR Spectra Using Bayesian Analysis, *Spectroscopy Letters*, 29(3), 393 – 400, 1996.
50. Harrington, P. B. and Isenhouer, T.L., Closure Effects in Infrared Spectral Library search Performance, *Appl. Spectrosc.*, 41, 1298, 1987.
51. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, NY, 1991.
52. Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Second Edition. Academic Press, Boston, 255 – 268, 1990.
53. Valafar, F., Valafar, H., and York, W.S., Identification of ^{1}H-NMR Spectra of Xyloglucan Oligosaccharide: A Comparative Study of Artificial Neural Networks and Bayesian Classification Using Nonparametric Density Estimation, *Int. Conf. Artif. Intelligence* 1999 (IC-AI'99), Las Vegas, NV, June 28 – July 1, 1999.
54. Rodgers, G.P., Dover, G.J., Noguchi, C.T., Schechter, A.N., Nienhuis, A.W., and Nienhuis, M.D., Hematologic Responses of Patients with Sickle Cell Disease to Treatment with Hydroxyurea, *New England J. Med.,* 322 Vol. 15, 1037 – 1044, April, 1990.
55. Charache, S., Terrin, M.L., Moore, R.D., Dover, G.J., Barton, F.B., Eckert, S.V., McMahon, R.P., and Bonds, D.R., Effect of Hydroxyurea on the Frequency of Painful Crises in Sickle Cell Anemia, *New England J. Med.,* 332, 1317 – 1322, May 18, 1995.
56. Charache, S., Dover, G.J., Moore, R.D., Eckert, S., Ballas, S.K., Koshy, M., Milner, PF., Orringer, E.P., Phillips, G. Jr., and Platt, O.S., Hydroxyurea: Effects on Hemoglobin F Production in Patients With Sickle Cell Anemia, *Blood*, 79(10), 2555 – 2565, May 15, 1992.
57. Powars, D.R., Weiss, J.N., Chan, L.S., and Schroeder, W.A., Is There a Threshold Level of Fetal Hemoglobin That Ameliorates Morbidity in Sickle Cell Anemia? *Blood*, 63(4), 921 – 926, April, 1984.
58. Charache, S., Barton, F.B., Moore, R.D., Terrin, M.L., Steinberg, M.H., Dover, G.J., Ballas, S.K., McMahon, R.P., Castro, O., and Orringer, E.P., Hydroxyurea and Sickle Cell Anemia. Clinical Utility of a Myelosuppressive "Switching" Agent. The Multicenter Study of Hydroxyurea in Sickle Cell Anemia, *Med.*, 75, Vol. 6, 300–325, November, 1996.
59. Steinberg, M.H., Lu, Z., Barton, F.B., Terrin, L.M., Charache, S., and Dover, G.J., Fetal Hemoglobin in Sickle Cell Anemia: Determinants of Response to Hydroxyurea, *Blood*, 89(3) 1078 – 1088, Feb, 1997.